使用 Mothur軟體 進行土壤微生物群落組成分析

農試所生技組 杜元凱 陳涵葳 林彥君

一、前言

土壤微生物有助於分解土壤中之有機物質、協助固氮作用進行及抑制作物病原菌增殖，土壤中之碳、氮、磷、硫的循環利用亦都需要微生物的參與，因此土壤微生物的多樣性在農業生態中特別重要。環境基因體學 (metagenomics) 是基於微生物學與次世代定序技術，直接萃取環境樣品DNA進行定序，藉由分析環境樣品中微生物組成之多樣性、豐富度，推估環境因子與微生物群落動態消長關係的研究 (Neelakanta and Sultana, 2013)。目前最常用的環境微生物分析方法有2種：其一針對微生物基因體之特定基因或區域 (如原核生物16S rRNA、真核生物18S rRNA及ITS)，以PCR增幅序列後搭配Illumina或其他定序平台做序列分析，可鑑別物種到屬層級，稱為擴增子定序 (amplicon sequencing)；另可選擇將樣本中所有微生物基因體進行定序分析，以求準確至種或菌株層級，稱為總體基因體定序 (whole metagenome sequencing, WMS)。以擴增子定序為例，細菌用於鑑別物種的16S rRNA基因全長約1.6 kb，具有V1－V9共9個高度變異區 (high variable regions)，而土壤樣本多選用V1－V3或V3－V4的變異區進行PCR增幅，並搭配Illumina平台進行定序。擴增子序列間的比對可依據97%之序列相似度分類為個別之操作分類單位 (operational taxonomic units, OTUs)，或是採用誤差校正模型講求高解析度的擴增序列變體 (amplicon sequence variants, ASVs) 找出更精確的基因體差異。隨著16S rRNA基因定序的普及，許多分析軟體相繼推出，譬如基於操作分類單位進行聚類分析的QIIME-uclust、MOTHUR及USEARCH-UPARSE，或是以擴增序列變體為主要分析流程的DADA2、Qiime2-Deblur、USEARCH-UNOISE3等套件。有鑑於不同分析軟體工具的流程差異，應根據實驗室既有資訊設備與訴求建立最

作　　者：杜元凱助理研究員
連絡電話：04-23317321

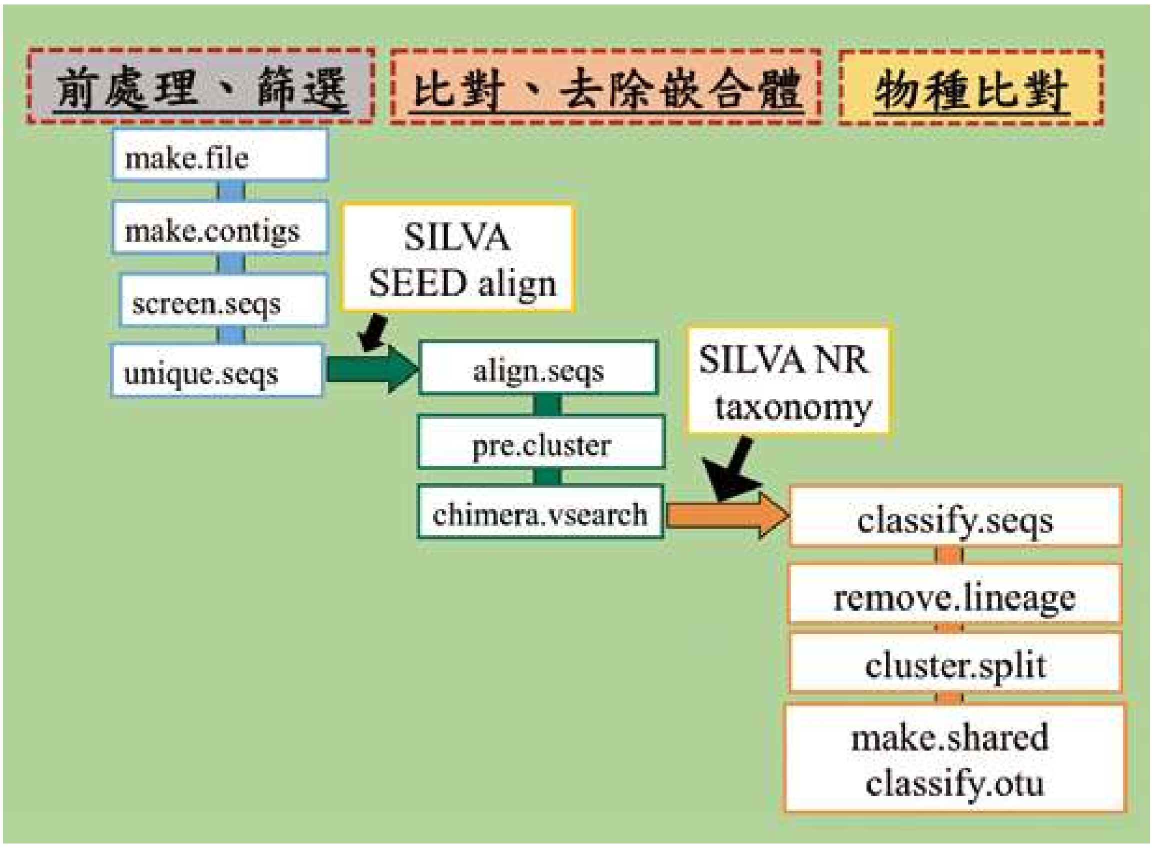
合理的分析流程。本文將簡述Mothur軟體之分析流程與指令涵義，提供對環境微生物資料統計與生物多樣性分析有興趣者在使用上之參考。

二、Mothur軟體介紹

Mothur軟體是2009年密西根大學醫學院的Patrick D. Schloss團隊以C++程式語言編寫所開發的分析工具，其整合了多個計算序列間成對距離矩陣之套件並優化其功能，透過輸入命令行完成序列組裝、過濾與篩選、資料比對及操作分類單位分析等步驟，亦可計算物種豐富度、均勻度等指標以分析樣本的微生物族群多樣性 (Schloss et al, 2009; Schloss, 2020)。Mothur結合功能強大的數據庫工具和系統發展套件包，做為免費的開源工具開放給大眾使用，使用者可於Mothur官方網站 (https://mothur.org/) 依據作業系統環境下載Windows、Mac OSX或Linux系統平台適合的壓縮檔，在解壓縮後即可使用，無須複雜的封包安裝。Mothur網站也提供SOP操作範例與教學論壇，方便使用者依據不同定序儀器處理序列資料微調樣本之分析流程 (Kozich et al, 2013)。

三、Mothur分析流程

Mothur軟體的分析流程包括序列資料分析、群落內的多樣性分析以及群落間多樣性分析，以下將逐一介紹(圖一)。



圖一、Mothur SOP**分析流程，包括樣本前處理**(pre-processing)**、組裝比對**(assembly)**、聚類與多樣性分析**(cluster analysis)。

(一) 序列資料分析

1. 序列讀取與組裝：make.file指令可設定檔案路徑、兩股序列檔案名稱及輸出名稱，再透過make.contigs指令將雙端定序資料合併成單一個fasta檔案。輸出檔案時可使用summary.seqs指令查看序列數量、reads的平均長度以及重複的同元序列 (homoploymer) 等資訊。
2. 序列資料篩選：合併後的片段重疊群 (contig) 可能含有相當多的非專一PCR擴增片段或低品質序列，都可經由screen.seqs指令設定篩選參數。例如使用341F/805R引子對增幅16S rRNA V3－V4目標片段，雙端測序合併後之序列長度估計為約430 bp，考慮到不同物種V3－V4序列鹼基可能含有插入或缺失，可設定maxlength=500去除過長的片段重疊群。另外若兩股鹼基配對不一致產生之不明確的序列 (ambiguous sequence)，可設定參數maxambig=0，提高分析的嚴謹度。若序列含有barcode和引子序列，亦可使用trim.seqs指令讀取指定的文件，刪除對應位置之序列資料。
3. 序列資料處理：為增加處理效能，序列資料可經由unique.seqs指令篩選單一序列 (unique sequence)，將序列中非保留性的特有區域擷取出來，以減少分析運算時間。產出的fasta檔案與經pcr.seqs指令過濾目標區域的參考基因體數據進行比對 (align.seqs)，再經由screen.seqs指令將因插入或刪除導致位置偏移的序列，重新訂定序列起始與終止位置。
4. 過濾與預聚類分析：使用filter.seqs辨識fasta檔案序列中的特定字符 (例如「-」、「.」、「N」等)，設定vertical=T及trump=.，移除欄位中不含序列資料的行數以節省檔案空間。在多數欄位被去除後，資料改以pre-cluster指令做歸類，根據參數diffs設定分群，以加快後續聚類分析計算，例如diffs=2代表序列將以出現2個或2個以內不同配對差異做分組，再按照序列出現次數由高至低排序。
5. 嵌合體去除：以指令chimera.vsearch搭配VSEARCH套件，可尋找含有嵌合體之序列。參數設定方面可在指令輸入時加入參考基因體如SILVA gold database，或使用denovo方法進行分析。執行完的輸出檔 (*.accnos) 可查閱被認定含有嵌合體的序列，再以remove.seqs指令從fasta檔案移除序列。
6. 物種比對：與參考資料庫中已知之物種比對，依據97%序列相似度的操作分類單位 (OTUs) 做分群。數據首先經classify.seqs指令進行分類演算法及資料庫比對，並經由remove.lineage指令移除葉綠體或粒線體基因等非目標序列。之後Mothur根據classify.seqs指令產出的*.taxonomy格式文件，使用dist.seqs與cluster指令進行成對距離計算與聚類分析，設定cutoff=0.03，將97%相同度之序列視為同一個OTU。

聚類分析亦可使用cluster.split替代原先的dist.seqs及cluster指令，該指令會將距離數值較大的參數分開個別運算再進行聚類分析，可加快運行速度並減少占用內存。處理完的*.list格式文件經由make.shared及classify.otu分別建立對應的OTU表單，make.shared指令產生*.shared格式文件以每個組別中每個OTU含多少序列做呈現，而classify.otu產生之*.taxonomy與*.tax.summary格式文件會記錄每個OTU所屬的物種階層。

(二) 群落內多樣性分析 (alpha diversity)

當樣本中物種數目越多，代表其物種豐度 (richness) 越大，可根據樣本中的序列數繪製成稀疏曲線 (rarefaction curve)，反應出樣本定序數量與觀察到的OTU數量的關係 (圖二)，或以豐度等級曲線 (rank-abundance curve) 同時呈現樣本中物種豐度與均勻程度。而Shannon-Wiener's Diversity Index和Simpson's Diversity Index可用於判別群落內多樣性的高低，前者指標同時考量物種數和物種均勻度，愈豐富均勻的群集其數值愈高，代表具較高的生物多樣性；後者則是以機率抽樣的概念隨機抽取兩個體來評估是否為同一物種，強調物種豐度的重要性，為方便呈現數值差異會以倒數函數之曲線關係作圖 (圖二)。

(三) 群落間多樣性分析 (beta diversity)

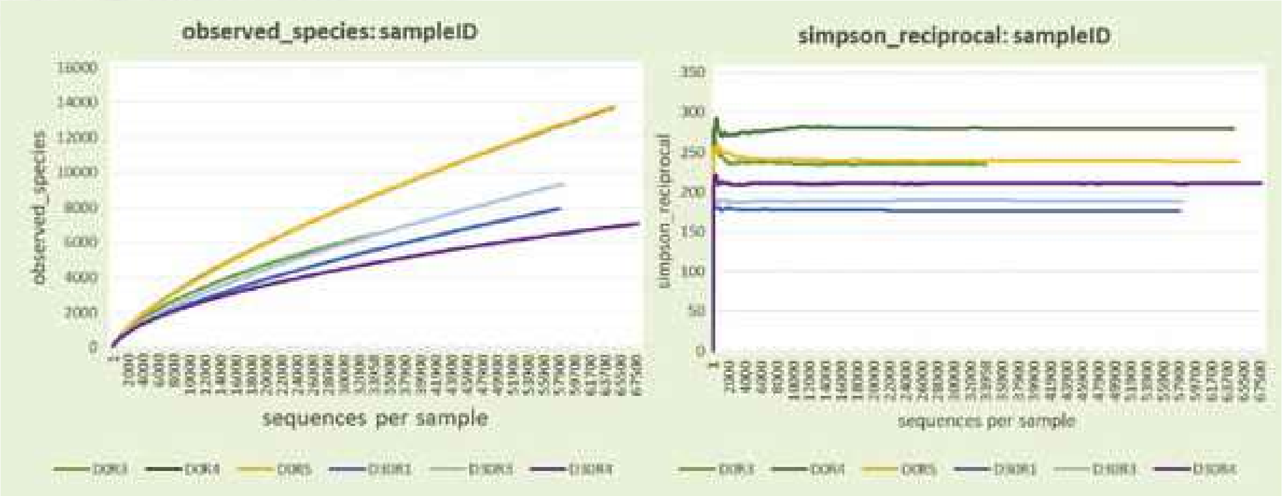
反應不同環境或處理介入前後微生物群落的差異，著重在環境因子與各群落組成差異。為了方便對照常以PCA、PCoA、heatmap以及階層式聚類法等方式將數據差異視覺化，使用的指標參數例如Jaccard index比較兩群落間共有物種數並推算相似性，Bray–Curtis similarity則計算兩群落物種數與組成是否具有一致性(圖三)。若要分析群落間物種的親緣關係，可另使用UniFrac套件 (Unique Fraction) 呈現系統發生樹，或利用Mothur內建的unifrac.weighted以及unifrac.unweighted兩種指令計算距離。

四、結語

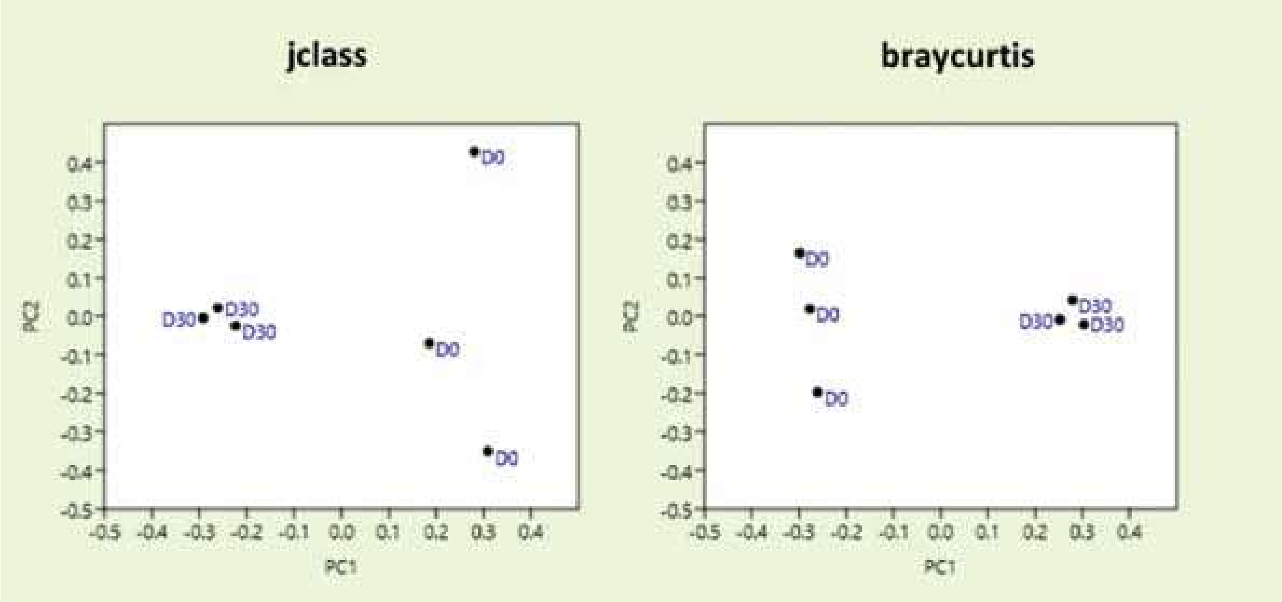
土壤中眾多微生物形成複雜的動態系統，隨著分子生物學的發展，利用次世代定序技術可幫助我們瞭解土壤微生物的種類及多樣性，以探討不同施肥、耕作方式對田間微生物群落的影響。以Mothur軟體套件分析土壤細菌之16S rRNA基因序列，可了解樣本中細菌群落之多樣性，推估微生物群落消長與環境因子之關聯性，若同時配合土壤理化性質、碳氮利用率、生化反應分析等資料項目，即可初步建構出環境微生物與土壤肥力的關聯性，未來可針對不同作物、品種及肥料處理，尋找出關鍵之指標微生物或更進一步窺知微生物所扮演之功能。

五、參考文獻

Kozich, J. J., Westcott, S. L., Baxter, N. T., Highlander, S. K., and Schloss, P. D. 2013 Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data



圖二、根據群落內的多樣性分析(alpha diversity)結果，繪製物種稀疏曲線(左)與Simpson's reciprocal index曲線(右)。



圖三、根據樣本間的生物多樣性(beta diversity)結果，分別依照參數Jaccard index(左)或Bray－Curtis similarity(右)計算兩群落一致性並繪製成XY散布圖。

on the MiSeq Illumina sequencing platform. Applied and Environmental Microbiology. 79:5112–5120.

Neelakanta, G., and Sultana, H. 2013 The use of metagenomic approaches to analyze changes in microbial communities. Microbiology Insights. 6:37–48.

Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., et al. 2009 Introducing mothur: open-source. platform-independent, community-supported software for describing and comparing microbial communities. Applied and Environmental Microbiology. 75:7537–7541.

Schloss, P.D. 2020 Reintroducing mothur: 10 Years Later. Applied and Environmental Microbiology. 86: e02343-19.